Hepatitis C virus genotype 4 (HCV-4) is very common in central Africa, prevalent in the Middle East, and is becoming increasingly frequent among southern Europeans. We have determined the complete nucleotide sequences of HCV-4f strains and investigated their phylogenetic relationships with other genotypes. We amplified the complete genome of two HCV subtype 4f strains, IFBT84 and IFBT88. The HCV-4f strains have a total of 9181 and 9304 nt, respectively, including the 5′ untranslated region followed by a single open reading frame. There was no evidence of genomic recombination in the IFBT84 and IFBT88 strains. The sequences of the HCV-4f strain genomes are closer to those of HCV-1b than to genotypes 2, 3, 5 and 6. It is important to know the full-length sequences of HCV-4 subtypes in order to classify them correctly and to obtain more detailed knowledge about HCV epidemiology and sensitivity to interferon.
The HCV-4f strains IFBT84 and IFBT88 were obtained from two chronically infected patients living in the Midipyrenees area of France. Serum sample IFBT84 was collected on 15 January 2002 and sample IFBT88 on 3 October 2001. They were genotyped by sequencing the NS5B region and underwent phylogenetic analysis in the laboratory of virology at Toulouse University Hospital (Sandres-Saune et al., 2003). The patients were women aged 57 and 25 years, respectively. The patient infected with IFBT84 was also infected with HIV-1. She was native of Cameroon and was treated unsuccessfully with interferon alpha (x-IFN) in 2002. The patient infected with IFBT88 was native of Gabon; she was infected with HCV alone and had had no HCV therapy.

HCV RNA was extracted from 140 μl serum using the QIAamp viral RNA method (Qiagen), 5 μl RNA was used to generate the cDNA. Two enzymes were used for reverse transcription, as recommended by the manufacturers: ThermoScript reverse transcriptase and SuperScript II RNase H− reverse transcriptase (Invitrogen). Reverse transcriptions were carried out using 2 pmol of an appropriate anti-sense PCR primer, 9183AS or 9300AS (Supplementary Table S1, available with the online version of this paper). RNA was denatured at 95 °C for 3 min, and the reaction mixtures were placed on ice for 3 min. Reverse transcription was started by adding reverse transcriptase and incubating at 43 °C for 1 h. This was followed by 20 cycles of 1 min at 53 °C, 1 min at 55 °C, followed by a final incubation at 70 °C for 15 min. The cycles were used to overcome secondary structures in the RNA templates that decreased the efficiency of reverse transcription.

Aliquots (5 μl) of cDNA were then used in the first PCR amplification in a final volume of 50 μl reaction mixture containing 0.5 mM dNTPs, 2.75 mM MgCl2, 3.75 U Expand Long Template enzyme mix (Roche) and 500 nM each outer primer (outer sense and outer anti-sense, Supplementary Table S1). Nested PCRs were carried out on 1 μl aliquots of the first amplification products in 50 μl with 0.5 mM dNTPs, 2.75 mM MgCl2, 3.75 U Expand Long Template enzyme mix (Roche) and 400 nM each inner primer (inner sense and inner anti-sense, Supplementary Table S1). The thermal profile of both PCR amplifications was 94 °C for 5 min, 10 cycles of 1 min at 94 °C, 1 min at the optimal annealing temperature of primers and 2–8 min at 68 °C for elongation depending on the fragment length, followed by 30 cycles using the same conditions as the former cycles with an extension time of 20 s for each successive cycle and a final elongation at 68 °C for 7 min. The PCR products were resolved on 0.8% agarose gels and visualized with ethidium bromide in 0.5 × TBE (Invitrogen).

PCR products of the predicted molecular size were purified (QIAquick PCR purification kit, Qiagen) and sequenced on both strands with the same primers as those used for amplification using Big Dye Terminator cycle technology on an Applied Biosystems ABI 3130xl Genetic Analyzer. Sequence analyses and DNA fragment assembly were performed using the Sequencher program (Gene Codes; Miller & Powel, 1994). Degenerate primers were designed after alignment with HCV subtype 4a sequence (GenBank accession no. Y11604). These primers, used for the first PCR amplification, produced the predicted size of PCR fragment. The first fragments obtained after the second round of PCR were used for sequencing and provided sequence information for preparing strain-specific primers. The subsequent PCRs employed the strain-specific primers at the near end of sequenced regions. The positions of each primer are shown relative to their position in the HCV-4a sequence Y11604. The nucleotide composition and ambiguous nucleotide bases were determined with the BioEdit sequence alignment editor (http://www.mbio.ncsu.edu/BioEdit/bioedit.html).

Alignment for phylogenetic analysis was carried out with CLUSTAL X 1.83 software (Thompson et al., 1994), the reproducibility of the branching pattern was determined using bootstrap analysis (1000 replicates) and the phylogenetic tree was drawn using TreeView 1.66 (Page, 1996) by the neighbour-joining (NJ) method. The pairwise genetic distances between complete sequences were calculated by the Tamura Nei+gamma correction method using Molecular Evolutionary Genetics Analysis software: MEGA (version 3.1) (Kumar et al., 2004). Continuous variables were tested using the Mann–Whitney test. P values <0.05 were considered to be significant. Recombination events were assessed using Bootscan and RDP algorithms implemented in Recombinant Detection Program version 2 (Martin et al., 2005). The GenBank accession numbers for sequences of the HCV-4f strains IFBT84 and IFBT88 are EF589160 and EF589161, respectively.

The HCV RNA of sample IFBT84 was reverse transcribed using ThermoScript reverse transcriptase. The Expand Long Template PCR system was used to amplify the entire nucleotide genome. To amplify the 5′ half of the HCV genome, the first round of PCR was carried out with outer primers KY80 and 5150AS, followed by nested PCRs using specific primers for strain IFBT84. Four HCV PCR fragments were amplified; the size of fragments was extended to 2733 bp. We could not amplify the 3′ half of the HCV genome, perhaps because there were secondary structures in the RNA templates that prevented PCR amplification. Reverse transcription using SuperScript II reverse transcriptase following the protocol described above provided reliable amplification. The first PCR amplification of the 3′ half of the HCV genome was performed with outer primers 4070S and 9183AS, followed by nested PCRs with specific primers (Supplementary Table S1). This produced eight HCV PCR fragments that extended the genomic sequence. This approach was used to amplify and sequence the complete genome of HCV-4f strain IFBT84 in two overlapping PCR fragments: the 5′ half was 5017 nt and the 3′ half was 4164 nt. The genome of HCV-4f strain IFBT84 has a total of 9181 nt; the 5′ untranslated region is 274 nt, followed by a single open reading frame.
reading frame of 8907 nt (2969 aa). The nucleotide composition is 21.23 mol% A, 30.03 mol% C, 27.58 mol% G, 21.14 mol% T and 0.02 mol% of mixed nucleotides (1 R and 1 Y).

The HCV cDNA of sample IFBT88 was synthesized with SuperScript II reverse transcriptase using 9183AS primer. The first round of PCR was carried out with outer primers KY80/5150AS and 4070S/9183AS, followed by nested PCRs using different primers (Supplementary Table S1); eleven HCV PCR fragments were amplified. The genome nucleotide sequence of the HCV strain from IFBT88 was further extended by reverse transcription with the 9300AS primer and followed by a first PCR amplification using outer primers 5498S and 9300AS followed by a hemi-nested PCR with inner primers 8690S and 9300AS. The combined overlapping PCR products nearly covered the complete genomic sequence (the 5' half was 5014 nt and the 3' half was 4291 nt). The complete genomic sequence is 9305 nt including the 5' untranslated region of 272 nt, followed by a single open reading frame of 9033 nt encoding 3011 aa. The nucleotide composition is 20.98 mol% A, 29.40 mol% C, 27.61 mol% G, 21.73 mol% T and 0.28 mol% of mixed nucleotides (10 Y, 9 R, 5 S, 1 K and 1 M). No insertions or deletions were noticed.

The phylogenetic tree, comparing 400 nt in the NS5b sequences of HCV strains IFBT84, IFBT88 (corresponding to nt 8177–8577) and controls representing different genotypes from GenBank, showed that these strains were HCV-4f subtypes (Fig. 1). This was confirmed with a phylogenetic tree comparing 211 nt of the E2 sequences of HCV strains IFBT84, IFBT88 (corresponding to nt 1330–1541) and HCV strains from Toulouse University Hospital (data not shown).

The genetic distance of complete nucleotide sequences of strains IFBT84 and IFBT88 was 0.092. No evidence of genomic recombination were found in either strain. Genetic distance calculations showed that the complete genome sequences from IFBT84 and IFBT88 strains were more closely related to sequences from type 4 (mean genetic distance: 0.176). Genotype 4 sequences were closer to genotype 1 than to other genotypes (P < 0.001) and subtype 4f and 4d sequences were closer to genotype 1 than subtype 4a sequences were (P < 0.03). Subtype 4f sequences were not significantly closer to genotype 1 than subtype 4d sequences were (P = 0.41). This was confirmed by a phylogenetic analysis with complete sequences from referenced strains representing various HCV genotypes and subtypes (Fig. 2). The analysis disclosed that HCV genotype 4 complete nucleotide sequences were clustered together with significant bootstrap values. Phylogenetic analysis (Fig. 2) showed that the complete genomes of HCV-4f IFBT84 and IFBT88 were closer to genotype 1 sequences than subtype 4a sequences were.

Although HCV genotype 4 strains are not common among the general population in France, they are occurring more frequently (Morice et al., 2001; Nicot et al., 2005), mainly among intravenous drug users (van Asten et al., 2004). The prevalence of HCV-4 infection was estimated at 7.4% in a large cohort of patients in the Midi-Pyrénées area of south-west France; 4% of the HCV-4 strains belonged to subtype 4f. Most of the patients infected with 4f strains in our area came from Africa, as did the two patients included in our study.

Patients infected with HCV genotype 4, like those infected with genotype 1, respond less well to HCV treatment than those infected with genotypes 2 and 3 (Legrand-Abravanel...
et al., 2005; Manns et al., 2001; Wohnsland et al., 2007; Zeuzem et al., 2004). In contrast, some recent studies have also shown that they are sensitive to PEG α-IFN + ribavirin, but the HCV subtype was not given (Basso et al., 2007; Fried et al., 2002; Hasan et al., 2004; Nguyen & Keeffe, 2005). The HCV-4f genome must be characterized before more knowledge can be obtained about this subtype, its epidemiology and its sensitivity to interferon. This can be useful for understanding why HCV genotype 4 infections respond differently to α-IFN-based therapies (Derbala et al., 2005).

The full-length sequences of the HCV strains IFBT84 and IFBT88 are, to the best of our knowledge, the first complete sequences of a subtype 4f genome, but we failed to characterize the extreme 5’ end of the genome as well as much of the 3’ untranslated region. These sequences can be used as references for classifying and confirming putative HCV-4f strains identified with partial sequences amplified from different genomic regions. Few sequences of 4a–4d HCV strain complete genomes have been determined; this is probably due to the extreme variability of HCV genotype 4; multiple subtypes of HCV genotype 4 have been isolated from distinct regions of central Africa, or from immigrants from these regions, indicating the restricted circulation of these subtypes and their limited transmission beyond these regions (Nguyen & Keeffe, 2005). Twelve subtypes have been identified in France by phylogenetic analysis of the NS5B region (Nicot et al., 2005). We have shown significant relationships between the sequences of complete genome of HCV subtype 4f, IFBT84 and IFBT88 genomes and sequences from type 4 by phylogenetic analysis and pairwise distance calculations. HCV-4 strains are more closely related to genotype 1 than to other genotypes and
subtype 4f is the genotype 4 closest to genotype 1. This relationship was recently demonstrated in a report describing two sequences of complete genomes of HCV strains belonging to subtypes 4d and 4a (Franco et al., 2007). Nevertheless, our report is possibly the first to demonstrate the significant relationship between HCV-4f and genotype 1, which could be used to understand how the HCV-1 and 4 subtypes have evolved. This relationship may be due to the HCV-1 and HCV-4 strains having a more recent common ancestor. HCV-4f may also have been one of the first genotype 4 strains to emerge while subsequent evolution led to the genetic diversification of both genotypes (Salemi & Vandamme, 2002). The patterns of HCV-1 and HCV-4 evolution could explain their similarly poor responses to antiviral therapy and their overlapping distributions in central west Africa. These two genotypes (Salemi & Vandamme, 2002). This study provides two complete nucleotide sequences of HCV-4f strains that can be used as references in the diagnosis of HCV-4f infections and in epidemiological studies. The data can also be used to better understand the intrinsic sensitivity of genotype 4 to α-IFN, and hence to generate efficient antiviral agents against this virus subtype.

References


Morice, Y., Roulot, D., Grando, V., Stirnemann, J., Gaul, E., Jeanlitis, V., Bentata, M., Jarrousse, B., Lortholary, O. & other authors (2001). Phylogenetic analyses confirm the high prevalence of hepatitis C virus (HCV) type 4 in the Seine-Saint-Denis district (France) and indicate seven different HCV-4 subtypes linked to two different epidemiological patterns. J Gen Virol 82, 1001–1012.


